Sublinear expectation linear regression

Lu Lin[*], Yufeng Shi, Xin Wang and Shuzhen Yang
Shandong University Qilu Securities Institute for Financial Studies
and School of Mathematics, Shandong University, Jinan, China

Abstract

Nonlinear expectation, including sublinear expectation as its special case, is a new and original framework of probability theory and has potential applications in some scientific fields, especially in finance risk measure and management. Under the nonlinear expectation framework, however, the related statistical models and statistical inferences have not yet been well established. The goal of this paper is to construct the sublinear expectation regression and investigate its statistical inference. First, a sublinear expectation linear regression is defined and its identifiability is given. Then, based on the representation theorem of sublinear expectation and the newly defined model, several parameter estimations and model predictions are suggested, the asymptotic normality of estimations and the mini-max property of predictions are obtained. Furthermore, new methods are developed to realize variable selection for high-dimensional model. Finally, simulation studies and a real-life example are carried out to illustrate the new models and methodologies. All notions and methodologies developed are essentially different from classical ones and can be thought of as a foundation for general nonlinear expectation statistics.

Key words: Sublinear expectation, G-normal, linear regression, model uncertainty, parameter estimation, variable selection.

Running head: Sublinear expectation regression.

[*] The corresponding author. Email: linlu@sdu.edu.cn. The research was supported by NNSF projects (11171188, 11071145, 11221061, and 11231005) and the 111 project (B12023) of China, NSF and SRRF projects (ZR2010AZ001 and BS2011SF006) of Shandong Province of China and K C Wong-HKBU Fellowship Programme for Mainland China Scholars 2010-11.

1 Introduction

Among all the assumption conditions imposed to classical statistical models, the most vital one may be that the model under study has a certain probability distribution that may or may not be known. The classical linear expectation and determinant statistics are built on the distribution-certainty or model-certainty. The distribution-certainty, however, is not always the case in practice, such as risk measure and super-hedging in finance. For related references see, e.g., El Karoui, Peng and Quenez (1997), Artzner, Delbaen, Eber and Heath (1999), Chen and Epstein (2002), Föllmer and Schied (2004). We also studied a relevant practical problem. It is known that in a financial market, non-performing loan (NPL) is always an important object to be monitored. The NPL ratio is of course related to some economic indicators such as loan-deposit ratio and capital adequacy ratio. We have used an indicator set and the corresponding data published in Vendors Database of China (2000-2010) to establish a regression relationship between the NPL ratio and the indicators in the set. It has been discovered that the regression error has a mean-uncertainty, meanly, the error mean is distributed in an interval $[-0.1833, 0.1747]$. We will discuss the issue in detail in Section 5.

Without distribution-certainty, the resulting expectation is nonlinear usually. The earlier works on nonlinear expectation may ascend to Huber (1981) in the sense of robust statistics or ascend to Walley (1991) in the sense of imprecise probabilities. In the recent decades, the theory and methodology of nonlinear expectation have been well developed and received much attention in some application fields such as finance risk measure and control. A typical example of the nonlinear expectation, called g-expectation (small g), was introduced in Peng (1997) in the framework of backward stochastic differential equations. As a further development, G-expectation (big g) and its related versions are proposed by Peng (2006). Under the nonlinear expectation framework, the most common distribution is the so-called G-normal distribution, which was first introduced in Peng (2006). Furthermore, as a theoretical basis of the nonlinear expectation, the law of large numbers as well as the central limit theorem were also established by Peng (2008 and 2009). Also, from different points of view, many authors studied nonlinear expectation, see, e.g., Denis and Martini (2006), Denis et al. (2011), Soner et al. (2011a, 2011b, 2012 and 2013). Other references include Chen and Peng (2000), Briand et al. (2000), Coquet et al. (2002), Gao (2009), Li and Peng (2011), Peng (1999, 2004, 2005 and 2009), Rosazza (2006), Song (2012), and Xu and Zhang (2009), among many others.

Contrary to the fast development of the nonlinear expectation in probability theory, little attention was paid to the related statistical models and statistical inferences to the best of our knowledge. Although the earlier work of Huber (1981) refers initially to a upper-lower expectation, a special nonlinear expectation, the main aspects focus on robust statistics and the underlying true model is supposed implicitly to have a certain distribution. Gross error model, for example, contains a certain true distribution in the contaminated distribution set, and based on such a distribution set, the supper-lower expectation can be defined; see, e.g., Strassen (1964) and Huber (1981). In classical statistical frameworks, the heteroscedastic model may be the closest one to the model-uncertainty aforementioned, but it only has variance-uncertainty and the corresponding inference methods do not involve any notion of nonlinear expectation. In nonparametric framework, the model structure is not specified, and in Bayesian framework, the model parameter is random. But the two statistical frameworks are essentially different from the model-uncertainty aforementioned and the corresponding methods are completely unrelated to any nonlinear expectation. In time series models, although the data depend on observation time, strict stationarity or weak stationarity is required to guarantee the certainty of statistical inferences. In a word, under the classical statistical framework, including parameter models, nonparametric models, Bayes models and time series models, the defined expectations are of linearity. Without this linearity, it is essentially difficult or impossible by using classical methods to achieve classical certain conclusions, such as the identifiability of model parameter, estimation consistency, asymptotic normality of the estimation and model selection consistency.

Under model-uncertainty frameworks, the classical statistics methods may no longer be available. The classical maximum likelihood, for example, is nonexistent or can not be uniquely determined due to without a certain likelihood function. Also the classical least squares estimation is invalid because the parameter is defined via linear expectation. Moreover, the classical statistical models such as linear regression models, may not be well-defined as their identifiability depends on mean-certainty; without mean-certainty, the regression function is unidentifiable. Furthermore, it will be verified by simulations in Section 5 that under the situation of model-uncertainty, usual methods may not work and even collapse nearly. Thus, to achieve the target of statistical inference, it is necessary to develop new statistical frameworks and new statistical methods.

The main contribution of our paper is to establish a framework of sublinear expectation regression for the model that has the distribution-uncertainty. Based on a sublinear expectation space, a sublinear expectation linear regression is defined and its identifiability is achieved. Our model is always available for the cases of variance-uncertainty and/or mean-uncertainty. Unlike classical regression, the new model tends to use a large value to predict the response variable and obtains the mini-max prediction risk. It implies that our method is a robust strategy and has potential applications in finance risk measure and management. Based on the representation theorem of sublinear expectation, new parameter estimation methods are suggested and the resulting estimators are asymptotically normal distributed for the case of high-frequency data. Finally, our method is extended to variable selection for high-dimensional regression. It is worth mentioning that under model-uncertainty framework, certainty-statistical inferences are established in this paper, including parameter-certainty, prediction-certainty and distribution-certainty of parameter estimation. The notions and methodologies developed here are nonclassical and original, and the theoretical framework establishes the foundations for general nonlinear expectation statistics.

The remainder of the paper is organized in the following way. In Section 2, a sublinear regression model is built and its identifiability is obtained. The estimation and prediction methods are suggested in Section 3. Also the asymptotic normality of estimators and the mini-max property of predictions are established in this section. The method is extended to variable selection for high-dimensional model in Section 4. Simulation studies and a real-life example are carried out in Section 5 to illustrate the new model and methodology. The proofs of the theorems and the definition of the sublinear expectation space are postponed to Appendix.

2 Sublinear expectation regression

In this section we establish a framework of sublinear expectation regression, including modeling, estimation, prediction and asymptotic properties.

2.1 Model

We consider the following linear regression model:

$$Y = \beta'\mathbf{x} + \varepsilon, \tag{2.1}$$

where Y is a scalar response variable, $\mathbf{x} = (X_1, \cdots, X_p)'$ is the associated p-dimensional covariate having a certain distribution $F_{\mathbf{x}}(x)$, and $\beta = (\beta_1, \cdots, \beta_p)'$ is a p-dimensional vector of unknown parameters. Furthermore, it is supposed that the error ε is independent of $\mathbf{x}$. We need the independence condition only for simplicity. The idea and methodology developed below can be extended to the dependent case, but the notations and algorithm are relatively complex. It is worth pointing out that the essential difference from the classical regression model is that here the error ε has distribution-uncertainty, which is defined in the following way.

Let Ω be a given set and $\mathcal{H}$ be a linear space of real valued functions defined on Ω. Furthermore, let $\mathbb{E}$ denote a sublinear expectation: $\mathcal{H} \to \mathbb{R}$, satisfying monotonicity, constant preserving, sub-additivity and positive homogeneity; for the details of the definitions see Appendix. The triple $(\Omega, \mathcal{H}, \mathbb{E})$ is then called a sublinear expectation space. In this paper, we assume that the random variable ε is defined on a sublinear expectation space $(\Omega, \mathcal{H}, \mathbb{E})$. It can be seen from the definition that the probability distribution of ε is uncertain. Under this situation, the independence between $\mathbf{x}$ and ε mentioned above is defined in the sublinear expectation space, which is a weak independence (2008 and 2009). For regression analysis, we suppose that $\mathcal{H}$ contains linear and quadratic functions, and although the sublinear expectation $\mathbb{E}$ is supposed to be existent, its exact form may be unknown. Thus, a remarkable point of view is that since regression analysis depends mainly on "expectation", we here only define a sublinear expectation space, instead of the well-accepted linear expectation.

By the representation theorem of sublinear expectation (Peng 2008 and 2009), the sublinear expectation of a function $g(\varepsilon) \in \mathcal{H}$ can be expressed as a supremum of linear expectations. Formally, there exists a family of linear expectations $\{E_f : f \in \mathcal{F}\}$ defined on $\mathcal{H}$ such that

$$\mathbb{E}[g(\varepsilon)] = \sup_{f\in\mathcal{F}} E_f[g(\varepsilon)] \quad \text{for } g \in \mathcal{H}, \tag{2.2}$$

and there exists a $f_g \in \mathcal{F}$ such that

$$\mathbb{E}[g(\varepsilon)] = E_{f_g}[g(\varepsilon)]. \tag{2.3}$$

Write

$$\overline{\mu} = \mathbb{E}[\varepsilon],\ \underline{\mu} = -\mathbb{E}[-\varepsilon],\ \overline{\sigma}^2 = \mathbb{E}[\varepsilon^2],\ \underline{\sigma}^2 = -\mathbb{E}[-\varepsilon^2].$$

Then, the intervals $[\underline{\mu}, \overline{\mu}]$ and $[\underline{\sigma}^2, \overline{\sigma}^2]$ characterize the mean-uncertainty and the variance-uncertainty of ε, respectively.

When $\mathbf{x}$ is a random variable, for regression modeling, it is necessary to clarify the sublinear expectation $\mathbb{E}[Y|\mathbf{x}]$ conditional on $\mathbf{x}$ since the nonlinear conditional expectation has not yet been defined in the existing literature. Actually, however, there is no obstacle to extend the nonlinear unconditional expectation to the nonlinear conditional expectation. By the representation theorem given above, for instance, the above $\mathbb{E}[Y|\mathbf{x}]$ can be defined as

$$\mathbb{E}[Y|\mathbf{x}] = \sup_{f_{Y|\mathbf{x}}\in\mathcal{F}_{Y|\mathbf{x}}} E_{f_{Y|\mathbf{x}}}[Y|\mathbf{x}],$$

where $\{E_{f_{Y|\mathbf{x}}} : f_{Y|\mathbf{x}} \in \mathcal{F}_{Y|\mathbf{x}}\}$ is a family of conditional linear expectations. With this definition, the properties of monotonicity, constant preserving, sub-additivity and positive homogeneity given in Appendix still hold.

Finally, we should note that it was assumed above that the covariate vector $\mathbf{x}$ has a certain distribution $F_\mathbf{x}$ and the intercept term of model (2.1) is zero. Here we need the distribution-certainty of $\mathbf{x}$ to guarantee that the regression coefficient vector β is identifiable; otherwise, when both ε and $\mathbf{x}$ do not have the distribution-certainty, β can not be uniquely determined. For details see Remark 2.1 below. The assumption on $\mathbf{x}$ and ε aforementioned is a practical condition. For example, if Y is a measure of a financial risk and $\mathbf{x}$ is the set of the corresponding economic indicators, then, usually the goal of regression analysis is to describe the risk measure Y for a given economic indicator set $\mathbf{x}$. Therefore, the indicator elements of $\mathbf{x}$ could be regarded as of distribution-certainty exactly or approximately. In this case, the model-uncertainty is derived from the unstable financial environments that can be grouped in the model error ε. On the other hand, we need the zero intercept to eliminate the estimation bias; without it, the estimation is inconsistent. For details see Remark 3.2 below.

2.2 G-normal regression

We first consider the case when the error ε is supposed to be G-normally distributed, namely,

$$\varepsilon \sim \mathcal{N} = N\left(\{0\} \times [\underline{\sigma}^2, \overline{\sigma}^2]\right). \tag{2.4}$$

Under this situation, ε has a certain zero mean but its variance is uncertain, a special distribution-uncertainty. As was defined by Peng (2006), ε is called G-normally distributed if it is defined on a sublinear expectation space $(\Omega, \mathcal{H}, \mathbb{E})$ and satisfies that for each $a, b \geq 0$,

$$a\varepsilon + b\,\bar{\varepsilon} \overset{d}{=} \sqrt{a^2 + b^2}\,\varepsilon,$$

where $\bar{\varepsilon}$ is an independent copy of ε and "$\overset{d}{=}$" stands for equal in distribution. For the definition and the representation of G-normal distribution see Peng (2006). It follows from the cash translatability of sublinear expectation given in Appendix that for regression model (2.1), if ε is G-normally distributed as in (2.4), then

$$\mathbb{E}[Y|\mathbf{x}] = \beta'\mathbf{x}. \tag{2.5}$$

The above relationship (2.5) could be thought of as a G-normal expectation regression because $\mathbb{E}$ is the G-normal expectation, a special sublinear expectation. Note that $\mathbf{x}$ has an identical distribution. Then, we have the following conclusion.

Proposition 2.1 *(1) If ε is G-normally distributed as in (2.4), then, the G-expectation of Y is identifiable in the sense that $\mathbb{E}(Y|\mathbf{x})$ can be uniquely determined by $\beta'\mathbf{x}$ as in (2.5). (2) Besides the condition above, if $E[\mathbf{x}\mathbf{x}']$ is a positive definite matrix, then, β is identifiable in the sense that β can be uniquely determined as*

$$\beta = (E[\mathbf{x}\mathbf{x}'])^{-1}E\{\mathbf{x}\mathbb{E}[Y|\mathbf{x}]\}, \tag{2.6}$$

where linear expectation E is taken under the certain distribution $F_{\mathbf{x}}(x)$.

The proof is given in Appendix. For the proposition, we have the following remark.

Remark 2.1

(1) The proposition implies that if the error ε is G-normally distributed and $\mathbf{x}$ has the distribution-certainty, then G-normal regression has both regression function-certainty and regression coefficient-certainty. The conclusion provides a theoretical basis for regression analysis such as parameter estimation and model prediction.

(2) From the proof of the proposition we can see that if $\mathbf{x}$ does not have the distribution-certainty but only a sublinear expectation is defined for $\mathbf{x}$, β can not be uniquely determined usually. Without the identifiability of β, there is no sense in modeling regression relationship.

(3) Here we emphasize the use of G-normal regression because a quadratic loss function will be employed below to construct a "quasi maximum likelihood" estimation; for details see the next section. In fact the notion proposed here can be directly extended to general mean-certainty

sublinear expectation regressions. Specifically, we only assume ε has the mean-certainty, instead of G-normal distribution. Under this situation, model (2.5) could be regarded as a mean-certainty sublinear expectation regression. With the point of view, the conclusions in Proposition 2.1 still hold.

2.3 Sublinear expectation regression

Now we investigate the model in which the error ε is mean-uncertain and variance-uncertain. By the cash translatability of the sublinear expectation given in Appendix, we have

$$\mathbb{E}[Y|\mathbf{x}] = \beta'\mathbf{x} + \overline{\mu}. \tag{2.7}$$

This model could be thought of as a sublinear expectation regression because $\mathbb{E}$ is a sublinear expectation. By (2.7) and similar arguments used in Proposition 2.1, we have the following conclusion.

Proposition 2.2 *(1) If $\underline{\mu} < \overline{\mu}$, then, given $\mathbf{x}$, the sublinear expectation of Y has a shift $\overline{\mu}$, more precisely, the sublinear expectation of Y has the framework of (2.7). (2) Besides the condition above, if $E[\mathbf{x}\mathbf{x}']$ is a positive definite matrix, then, β is identifiable, more precisely, β can be uniquely expressed by*

$$\beta = (E[\mathbf{x}\mathbf{x}'])^{-1}E\{\mathbf{x}\mathbb{E}[Y|\mathbf{x}]\} - \overline{\mu}(E[\mathbf{x}\mathbf{x}'])^{-1}E[\mathbf{x}]. \tag{2.8}$$

Particularly, if $E[\mathbf{x}] = 0$, then

$$\beta = (E[\mathbf{x}\mathbf{x}'])^{-1}E\{\mathbf{x}\mathbb{E}[Y|\mathbf{x}]\}. \tag{2.9}$$

The proof is also presented in Appendix. From the proposition, we have the following findings.

Remark 2.2

(1) The conclusions in the proposition are somewhat surprising because they suggest a nonclassical point of view and provide a methodological development. That is to say, in the face of mean-uncertainty, we can still uniquely determine the parameter vector β and then use the mean-shift framework $\beta'\mathbf{x}+\overline{\mu}$, instead of $\beta'\mathbf{x}$, to predict the response variable Y. Such a framework reflects the robust feature of sublinear expectation regression. If Y is a measure of the risk of a financial product, then the sublinear expectation regression tends to use a relatively large value to predict risk and moreover, and the increment of risk measure is just the sublinear expectation $\overline{\mu}$ of the error ε.

(2) It is worth mentioning that when the model does not have the mean-certainty, the representation (2.8) of regression coefficient vector β is different from the representation in (2.6) for the mean-certainty model, in other words, the representation (2.6) in the mean-certainty model is a special case of the representation (2.8) with $\overline{\mu} = 0$. This is an essential feature of sublinear expectation regression, i.e., in the mean-uncertainty framework, the regression coefficient vector β depends on the nonlinear expectation of error ε. Such a feature is totally different from classical linear expectation regression because in the linear expectation regression framework, the regression coefficient vector β has an error-free representation as

$$\beta = (E[\mathbf{x}\mathbf{x}'])^{-1}E[\mathbf{x}Y].$$

On the other hand, when $E[\mathbf{x}] = 0$, the parameter representation in (2.9) is free of $\overline{\mu}$. In the following, we mainly focus on the parameter representation in (2.9) because we will see that without $\overline{\mu}$, the corresponding estimator of β is relatively simple and is asymptotically unbiased.

3 Estimation and prediction

It is supposed in this section that the dimension p of β is fixed. Let $\{(Y_i, \mathbf{x}_i : i = 1, \cdots, N\}$ be a sample from model (2.1), satisfying

$$Y_i = \beta'\mathbf{x}_i + \varepsilon_i, \ i = 1, \cdots, N. \tag{3.1}$$

Unlike the classical ones, here $Y_1, \cdots, Y_N$ may have distribution-uncertainty due to the distribution-uncertainty of $\varepsilon_1, \cdots, \varepsilon_N$. Then the corresponding estimation method should be different from the classical ones that only apply to linear expectation regression models.

It seems that we can use (2.9) to construct the estimator of β as it presents a closed expression for β. However, in the expression, $\mathbb{E}[Y|\mathbf{x}]$ is a sublinear conditional expectation, like the classical ones, its estimation does involve multivariate nonparametric methods and therefore faces the curse of dimensionality if the dimension p of $\mathbf{x}$ is high. To avoid the problem, we now introduce a mini-max method to construct the estimator of β.

Case 1. We first consider the case of ε having the mean-certainty. Because Y has the sublinear expectation $\beta'\mathbf{x}$ given $\mathbf{x}$, theoretically, we should choose β so that it can minimize the sublinear expectation loss:

$$\mathbb{E}\left[(Y - \beta'\mathbf{x})^2\right]. \tag{3.2}$$

We can easily verify that the above sublinear expectation loss is a convex function function of β. Thus the optimization problem has a unique global optimal solution. The above is in fact a sublinear expectation least squares. It is worth mentioning that under G-normal distribution, we have that if φ is a convex function, then

$$\mathbb{E}[\varphi(\varepsilon)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi(\overline{\sigma}u) \exp\left\{-\frac{u^2}{2\overline{\sigma}^2}\right\} du,$$

and if φ is a concave function, then

$$\mathbb{E}[\varphi(\varepsilon)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi(\underline{\sigma}u) \exp\left\{-\frac{u^2}{2\underline{\sigma}^2}\right\} du.$$

For details refer to Peng (2006). These imply that under the convex function and concave function spaces, the G-normal has density functions $\frac{1}{\sqrt{2\pi}\overline{\sigma}} \exp\left\{-\frac{u^2}{2\overline{\sigma}^2}\right\}$ and $\frac{1}{\sqrt{2\pi}\underline{\sigma}} \exp\left\{-\frac{u^2}{2\underline{\sigma}^2}\right\}$, respectively. Therefore, the above sublinear expectation least squares could be thought of as a "quasi maximum likelihood".

To implement the estimation procedure, we need the following assumption:

C1. There exists an index decomposition: $I_i, i = 1, \cdots, m$, such that when $(ij) \in I_i$, $\varepsilon_{i1}, \cdots, \varepsilon_{in_i}$ are independent and have an identical distribution.

This condition is essentially implied in the conclusion (2.2) of the representation theorem given in Subsection 2.1. Thus, m should be equal to the number of functions in $\mathcal{F}$ if $\mathcal{F}$ only contains finite number of functions; otherwise, m should tend to infinity and in this case, the condition *C1* is only an approximation of the true one. We will further weaken *C1* and suggest a data-driven decomposition after Theorem 3.1 given below. From now on we suppose that the numbers of elements in $I_i, i = 1, \cdots, m$, are equal, i.e., $n_1 = n_2 \cdots = n_m = n$, without loss of generality. Because it is assumed that $\varepsilon_{i1}, \cdots, \varepsilon_{in}$, are identically distributed, the independence in condition *C1* is the same as that in linear expectation framework, instead of the independence in the nonlinear expectation. Here we need independence only for simplicity. Without the independence assumption, for example, $\varepsilon_{i1}, \cdots, \varepsilon_{in}$ are weakly dependent, the conclusions given below still hold; for weakly dependent processes and the properties of estimation see for example Rosenblatt (1956, 1970), Kolmogorov and Rozanov (1960), Bradley and Bryc (1985), and Lu and Lin (1997). Furthermore, a common decomposition is built according to the observation time order, more precisely, $\varepsilon_1, \cdots, \varepsilon_N$ are reindexed as $\varepsilon_{ij} = \varepsilon_{(i-1)n+j}, i = 1, \cdots, m, j = 1, \cdots, n$, and then the index sets I_i's are defined as $I_i = \{(ij) : j = 1, \cdots, n\}$. It is known that in a small time interval, the characteristic of data could be regarded as to be changeless exactly or approximately. Under this point of view, condition *C1* is relatively mild. Also we can decompose the index set according to the values of Y in a descending order for example.

Denote by F_i the common distribution function of $\varepsilon_{ij}, (ij) \in I_i$. According to the representation theorem of sublinear expectation given in (2.2), sublinear expectation loss (3.2) can be written as $\max\limits_{1\le i\le m} E_{F_i}[(Y-\beta'\mathbf{x})^2]$ and therefore its empirical version is

$$\max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n} \left[Y_{ij}-\beta'\mathbf{x}_{ij}\right]^2. \tag{3.3}$$

By minimizing (3.3), we obtain a mini-max estimator of β as

$$\hat\beta_G = \arg\min_{\beta\in\mathcal{B}} \max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n} \left[Y_{ij}-\beta'\mathbf{x}_{ij}\right]^2. \tag{3.4}$$

It can be easily verified that $\max\limits_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n} \left[Y_{ij}-\beta'\mathbf{x}_{ij}\right]^2$ is a convex function function of β. Thus the resulting estimator $\hat\beta_G$ is a unique global optimal solution in the above optimization problem. Furthermore, such an estimation procedure can be easily implemented via, for example, genetic algorithm. Denote $\sigma_i^2 = E(\varepsilon_{ij}^2)$ for $(ij) \in I_i$ and $\sigma_{i_*}^2 = \max\limits_{1\le i\le m} \sigma_i^2$, and for simplicity, assume that

$$\sigma_{i_*}^2 > \sigma_i^2 \text{ for all } i \ne i_*.$$

The mini-max estimator above is asymptotically normally distributed. The following theorem gives the details.

Theorem 3.1 *For the mean-certainty model, if condition C1 holds and* $E[\mathbf{x}\mathbf{x}']$ *is a positive definite matrix and* $n\to\infty$ *as* $N\to\infty$*, then*

$$\sqrt{n}(\hat\beta_G-\beta) \xrightarrow{d} N\left(0, \sigma_{i_*}^2(E[\mathbf{x}\mathbf{x}'])^{-1}\right) \quad (N\to\infty),$$

where $\xrightarrow{d}$ *stands for convergence in distribution and* $N\left(0, \sigma_{i_*}^2(E[\mathbf{x}\mathbf{x}'])^{-1}\right)$ *is a classical normal distribution.*

This theorem establishes the theoretical foundation for further statistical inferences such as constructing confidence intervals and test statistics. From the proof of the theorem given in Appendix we can see that condition *C1* can be replaced by the following relatively weak condition:

C1'. $\varepsilon_{i_*1}, \cdots, \varepsilon_{i_*n}$ are independent and have an identical distribution.

This condition only involves the errors with indexes in I_{i_*}. Thus it is relatively common and is implied in (2.3), the second conclusion of the representation theorem. However, recognizing the fact that the number n of data in each small time slice I_i should be relative large, condition *C1* or *C1'* only applies to the case of high-frequency data. Moreover, by the two conditions, it is implicitly assumed that the index compositions $I_i, i = 1, \cdots, m$, or I_{i_*} are known completely. Under some situations, however, it is difficult or impossible to get such exact compositions in advance. Thus, data-driven decompositions are desired in practice. Now we briefly discuss this issue. By condition *C1'*, the proof of Theorem 3.1 and (2.3), the mini-max estimator in (3.4) can be approximately recasted as

$$\hat{\beta}_G = \arg\min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{j=1}^{n} \left[Y_{i_*j} - \beta' \mathbf{x}_{i_*j}\right]^2. \tag{3.5}$$

Thus, a simple approach is to identify I_{i_*} or its subset. Let $I_i^0 = \{(ij) : j = 1, \cdots, n^0\}$, $i = 1, \cdots, m^0$, be the initial compositions according to the observation time order for example, where $n^0 > p$. Note that in the case of mean-certainty, the common LS estimator $\hat{\beta}_{LS}$ of β is consistent. We then arrange $\sum_{j=1}^{n^0}(Y_{ij} - \hat{\beta}'_{LS}\mathbf{x}_{ij})^2, i = 1, \cdots, m^0$, in the descending order as

$$\sum_{j=1}^{n^0}(Y_{i_1j} - \hat{\beta}'_{LS}\mathbf{x}_{i_1j})^2 \geq \sum_{j=1}^{n^0}(Y_{i_2j} - \hat{\beta}'_{LS}\mathbf{x}_{i_2j})^2 \geq \cdots \geq \sum_{j=1}^{n^0}(Y_{i_{m^0}j} - \hat{\beta}'_{LS}\mathbf{x}_{i_{m^0}j})^2.$$

From (2.3) we can see that when n^0 is relatively small, the index set $I_{i_1}^0 = \{(i_1j) : j = 1, \cdots, n^0\}$ can be chosen as an initial choice of I_{i_*} or a subset of

I_{i_*}. We then use the data in $I^0_{i_1}$, together with approximate formula (3.5), to build the estimator. Since the data size in $I^0_{i_1}$ may be small, it is necessary to enlarge the initial choice $I^0_{i_1}$. To this end, we consider the following hypothesis test:

$$H_0 : \sigma_2^2 = v_1^2 \;\Leftrightarrow\; H_1 : \sigma_2^2 < v_1^2,$$

where σ_2^2 is the supposed variance of $\varepsilon_{i_2 j}$ for $(i_2 j) \in I^0_{i_2} = \{(i_2 j) : j = 1, \cdots, n^0\}$ and $v_1^2 = \frac{1}{n^0 - p}\sum_{j=1}^{n^0}(Y_{i_1 j} - \hat{\beta}'_{LS}\mathbf{x}_{i_1 j})^2$. Classical methods can used to test the hypothesis H_0. If H_0 is not rejected, then $I^0_{i_1} \bigcup I^0_{i_2}$ could be chosen as an enlarged choice of I_{i_*}. The procedure is repeated until the remainder variances are significantly smaller than v_1^2. Also we can use cluster analysis and/or discriminant analysis to achieve this goal.

After the estimator $\hat{\beta}_G$ is obtained, a natural prediction of Y is

$$\hat{Y} = \hat{\beta}'_G \mathbf{x}. \tag{3.6}$$

If model-uncertainty is ignored and common least squares (LS) method is used to construct the estimator $\hat{\beta}_{LS}$ of β, then the LS-based prediction is

$$\hat{Y} = \hat{\beta}'_{LS} \mathbf{x}. \tag{3.7}$$

Comparing the two estimators by maximum prediction risk and average prediction risk, we have the following conclusion.

Theorem 3.2 *Under the condition of the mean-certainty, whether the variance-uncertainty exists or not, the following relationships always hold:*

$$\max_{1\le i\le m}\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \hat{\beta}'_G\mathbf{x}_{ij}\right]^2 \le \max_{1\le i\le m}\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \hat{\beta}'_{LS}\mathbf{x}_{ij}\right]^2,$$

$$\frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \hat{\beta}'_G\mathbf{x}_{ij}\right]^2 \ge \frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \hat{\beta}'_{LS}\mathbf{x}_{ij}\right]^2.$$

From the theorem, we have the following finding.

Remark 3.1

The theorem indicates that sublinear expectation regression is a robust strategy that can reduce maximum prediction risk. Thus, it can be expected that such a regression could be useful for measuring and controlling financial risks.

Case 2. We now consider the case of ε having both the mean-uncertainty and the variance-uncertainty. In this case Y has the sublinear expectation $\beta'\mathbf{x}+\overline{\mu}$ given $\mathbf{x}$. Theoretically, we should choose β so that it can minimize the sublinear expectation loss:

$$\mathbb{E}\left[(Y-\beta'\mathbf{x}-\overline{\mu})^2\right]. \tag{3.8}$$

However, we cannot directly implement the estimation procedure as $\overline{\mu}$ is unknown usually. We thus design a profile estimation procedure as follows. Let $\hat{\beta}$ be an initial estimator of β, which may be the estimator obtained in Case 1 or by common least squares. When $E[\mathbf{x}]=\mathbf{0}$, Proposition 2.1 and Proposition 2.2 show that the regression coefficient vectors in Case 1 and Case 2 are equal to each other and thus such an initial estimator is also consistent for Case 2. We then estimate $\overline{\mu}$ by

$$\hat{\overline{\mu}}=\max_{1\le i\le m}\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij}-\hat{\beta}'\mathbf{x}_{ij}\right]$$

and finally estimate β by

$$\tilde{\beta}_G=\arg\min_{\beta\in\mathcal{B}}\max_{1\le i\le m}\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij}-\beta'\mathbf{x}_{ij}-\hat{\overline{\mu}}\right]^2. \tag{3.9}$$

Denote $\mu_i=E[\varepsilon_{ij}]$, $\sigma_i^2=E(\varepsilon_{ij}-\mu_i)^2$, $v_i^2=\sigma_i^2+(\overline{\mu}-\mu_i)^2$ and $v_{k_*}^2=\max\{v_i^2: i=1,\cdots,m\}$, and for simplicity, assume $v_{k_*}^2>v_i^2$ for all $i\neq k_*$. By the same argument as that in Theorem 3.1, we can prove that the estimator $\tilde{\beta}_G$ is asymptotically normal distributed. The following theorem presents the details.

Theorem 3.3 *For mean-variance-uncertainty, if condition C1 holds,* $E[\mathbf{x}] = \mathbf{0}$ *and* $E[\mathbf{x}\mathbf{x}']$ *is a positive definite matrix and* $n \to \infty$ *as* $N \to \infty$*, then*

$$\sqrt{n}(\tilde{\beta}_G - \beta) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2_{k_*}(E[\mathbf{x}\mathbf{x}'])^{-1}\right) \quad (N \to \infty).$$

For proof of the theorem see Appendix. This theorem establishes a foundation for further statistical inferences and data analyses. Here we also need to check the condition *C1*. From the estimation procedure given above, we see that it is asymptotically equivalent to determine two index sets, in which the mean of the error and $\frac{1}{n}\sum_{j=1}^{n}(Y_{k_*j} - \beta'\mathbf{x}_{k_*j} - \overline{\mu})^2$ achieve the maximum values $\overline{\mu}$ and $v^2_{k_*}$, respectively. The approaches are similar to those used in Case 1 and thus details are omitted here. On the other hand, it is worth pointing out that under the situation of mean-certainty, the condition $E[\mathbf{x}] = 0$ is vital for estimation consistency. The following remark will explain its importance.

Remark 3.2

For a model that has the mean-variance-uncertainty, if $E[\mathbf{x}] \neq 0$, then, by the relationship between (2.8) and (2.9), we can prove the estimator $\hat{\overline{\mu}}$ of $\overline{\mu}$ has an asymptotic bias: $-\overline{\mu}E[\mathbf{x}'](E[\mathbf{x}\mathbf{x}'])^{-1}E[\mathbf{x}]$. As a result, if $E[\mathbf{x}] \neq 0$, by the same argument as that used in the proof of Theorem 3.3, it can be verified that the estimator $\tilde{\beta}_G$ has an asymptotic bias as

$$\text{bias}(\tilde{\beta}_G) = (c\overline{\mu} - \mu_{k_*})(E[\mathbf{x}\mathbf{x}'])^{-1}E[\mathbf{x}],$$

where $c = 1 - E[\mathbf{x}'](E[\mathbf{x}\mathbf{x}'])^{-1}E[\mathbf{x}]$. Furthermore, without $E[\mathbf{x}] = 0$, the bias-correction is essentially difficult because, under the model-uncertainty framework, the law of large numbers can not strictly determine the consistency of sample mean; see Peng (2007 and 2008). On the other hand, the condition $E[\mathbf{x}] = 0$ induces that the intercept term in model (2.1) should be zero, which implies that if the intercept is nonzero, the estimation bias can not be completely eliminated and thus the estimator is inconsistent.

With the estimator, a natural prediction of Y is

$$\tilde{Y} = \tilde{\beta}'_G \mathbf{x} + \hat{\overline{\mu}}. \tag{3.10}$$

Similar to the properties in Theorem 3.2, the prediction $\tilde{Y}$ can obtain the mini-max prediction risk.

Theorem 3.4 *Whether or not the mean-uncertainty and the variance-uncertainty exist, the following relationship always holds:*

$$\max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \tilde{\beta}'_G \mathbf{x}_{ij} - \hat{\overline{\mu}}\right]^2 \le \max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \hat{\beta}'_{LS} \mathbf{x}_{ij}\right]^2.$$

It shows that our proposal is a robust strategy and is therefore useful for measuring and controlling financial risk. Meanwhile, the simulation study given in Section 5 will verify that when model has mean-variance-uncertainty, the average prediction error of the new method is usually smaller that of the LS method, namely,

$$\frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \tilde{\beta}'_G \mathbf{x}_{ij} - \hat{\overline{\mu}}\right]^2 < \frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \hat{\beta}'_{LS} \mathbf{x}_{ij}\right]^2.$$

It is because the prediction bias of $\hat{\beta}'_{LS}\mathbf{x}$ is between $\underline{\mu}$ and $\overline{\mu}$, which is not ignorable, especial for the case of $\underline{\mu}\,\overline{\mu} > 0$.

4 Variable selection

In this section we focus on the case when the dimension $p = p_N$ tends to infinity as sample size N increases. Under this situation, model (2.1) is further supposed to be sparse in the sense that only d components $\beta_{l_k}, k =$

$1, \cdots, d$, are nonzero with $d \ll N$. Without loss of generality, it is assumed that the first d coefficients $\beta_1, \cdots, \beta_d$ are nonzero.

Note that under sublinear expectation framework, the identifiability theory about β and $\mathbb{E}[Y|\mathbf{x}]$ given in Proposition 2.1 and Proposition 2.2 is free of the dimension p. Thus, for high-dimensional model, the conclusions in Proposition 2.1 and Proposition 2.2 still hold. With the identifiability, we can investigate variable selection, parameter estimation and model prediction under sublinear expectation framework. For simplicity, we only use the LASSO (Tibshirani (1996) and Zou (2006)) to achieve our goals. The method developed below can be extended to other penalty methods such as SCAD (Fan and Li (2001) Fan and Peng (2004)) and Dantzig selector (Candés and Tao (2007)).

We first consider the case of ε having the mean-certainty. The theoretical objective function is defined as

$$\mathbb{E}\left[(Y - \beta'\mathbf{x})^2\right] + \lambda \sum_{k=1}^{p} |\beta_k|, \tag{4.1}$$

where $\lambda \geq 0$ is a tuning parameter, which controls the amount of regularization applied to the estimate. Under condition *C1*, the empirical version of the above objective function is

$$\max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^{n} \left[Y_{ij} - \beta'\mathbf{x}_{ij}\right]^2 + \lambda \sum_{k=1}^{p} |\beta_k|. \tag{4.2}$$

By minimizing (4.2), we can achieve the goals of variable selection and parameter estimation simultaneously. It can be verified easily that the objective function above is a convex function of β. Then, the global minimum solution exist uniquely. Furthermore, such an optimization procedure can be easily implemented via, for example, genetic algorithm. Denoted by $\hat{\beta}_G$ the solution of the optimization problem (4.2). Note that most components of $\hat{\beta}_G$ are shrunk to zero by choosing a suitable tuning parameter λ. Then, the goal of variable selection can be realized. After variable selection and parameter estimation being completed, a natural prediction of Y can be chosen as

$$\hat{Y} = \hat{\beta}'_G \mathbf{x}. \tag{4.3}$$

Similar to the arguments in Theorem 3.2, our method is a robust strategy because the selected model can reduce the maximum prediction risk. Thus, the selected model by sublinear expectation can be employed to measure and control financial risks.

From the proof of Theorem 3.1, we see that when n is large enough, the term of order $O_p(1/n)$ can be ignored and the objective function above is approximately equal to

$$\frac{1}{n}\sum_{j=1}^{n}\left[Y_{i_*j} - \beta'\mathbf{x}_{i_*j}\right]^2 + \lambda\sum_{k=1}^{p}|\beta_k|, \tag{4.4}$$

where i_* is the index of the interval I_{i_*} in which the variance of ε achieves the maximum value. This representation implies that the properties of variable selection and parameter estimation, such as the selection consistency and the Oracle property of the estimator, are the same as those of the standard LASSO. So it is unnecessary to restudy these theoretical properties under the sublinear expectation framework. However, this representation shows that the number of data in each small time slice I_i should be relative large. Therefore our method only applies to high-frequency data.

If ε possesses both the mean-uncertainty and the variance-uncertainty, as was shown in the previous selection, we need the condition $E[\mathbf{x}] = 0$ to guarantee the consistency of estimation. Variable selection and parameter estimation can be obtained by minimizing the following objective function

$$\max_{1\le i\le m}\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \beta'\mathbf{x}_{ij} - \hat{\overline{\mu}}\right]^2 + \lambda\sum_{k=1}^{p}|\beta_k|. \tag{4.5}$$

Here $\hat{\overline{\mu}}$ is an initial estimator of $\overline{\mu}$ defined by

$$\hat{\overline{\mu}} = \max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \hat{\beta}_G'\mathbf{x}_{ij}\right],$$

where $\hat{\beta}_G$ may be the solution by minimizing (4.2). Denote by $\tilde{\beta}_G$ the corresponding solution. Then a prediction of Y is chosen as

$$\tilde{Y} = \tilde{\beta}_G'\mathbf{x} + \hat{\overline{\mu}}. \tag{4.6}$$

Also this prediction achieves the mini-max prediction risk and the prediction value tends to be larger.

5 Simulation study and real data analysis

5.1 Simulation study

In this section we present several simulation examples to compare the finite sample performances of the sublinear expectation regression proposed in this paper with the existing competitors, such as the classical LS regression and the LASSO regression. To get comprehensive comparisons, we use the mean square error (MSE), maximum prediction error (MPE) and average prediction error (APE), together with scatter plots of the estimation and prediction, to assess the different methods. From the simulations given below, we will get the following findings: (1) The new methods can significantly reduce the MPE under all the situations; (2) When the model has the mean-certainty, the advantages of the new methods over the classical LS methods are not very obvious; (3) For the case of the mean-uncertainty, the predictions of the classical LS methods do not work and even collapse nearly, but the new methods can get a valid prediction because the impact of the mean-uncertainty on the new methods can be successfully eliminated by the use of the sublinear expectation of the error. Thus, our proposals are robust to the uncertainties of mean and variance and particularly, for the case of the mean-uncertainty, the advantages of ours are rather obvious.

Experiment 1. We first consider the following simple linear model

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_2 X_2 + \varepsilon.$$

In the simulation procedure, the regression coefficients are chosen as $\beta_k = 1, k = 1, 2, 3$, the observation values of X_k are independent and identically distributed from $N(10, 2), k = 1, 2, 3$. We choose $\varepsilon \sim N(\{0\} \times [0, 3])$, a G-normal distribution with certain zero mean. In this case, the model has the mean-certainty. The following way is used to generate the data of G-normal distribution approximately. Firstly, generate variance values $\sigma_i^2, i = 1, \cdots, m$, from the uniform distribution $U[0, 3]$, and then generate the values $\varepsilon_{ij}, j = 1, \cdots, n$, of ε from the common normal distribution $N(0, \sigma_i^2)$. For $m = 10$ and $n = 10$, the simulation results are reported in Table 1, in which MSE, MPE and APE denote the mean squared error, maximum prediction error and average prediction respectively; for the definitions of MPE and APE see Proposition 2.2. It is clear by the simulation results that the MSE and APE of common LS estimation $\hat{\beta}_{LS}$ are significantly smaller than those of the G-normal estimation $\hat{\beta}_G$. Such a result is not surprising because, under the mean-certainty model, the common LS estimation $\hat{\beta}_{LS}$ is consistent but the construction of the new estimation $\hat{\beta}_G$ only uses the data in a small time interval (essentially, the number of the data used to construct the estimator $\hat{\beta}_G$ is only 10). On the other hand, the MPE by the new one $\hat{\beta}_G$ is significantly small than that by the LS estimator $\hat{\beta}_{LS}$, which implies than the new method can reduce the maximum prediction risk and therefore is a robust strategy.

The simulation results above indicate that when model has the mean-certainty, the advantages of the new methods over the common LS are not rather obvious. Moreover, the new methods even have the disadvantage of instability. In the following, we will see that when model has the mean-

Table 1: Simulation results of estimation and prediction for Experiment 1 with $m = 10$ and $n = 10$

MSE	β_1	β_2	β_2	MPE	APE
$\hat{\beta}_G$	0.0080	0.0301	0.0315	6.0259	3.5388
$\hat{\beta}_{LS}$	0.0026	0.0045	0.0037	6.6122	2.8584

uncertainty, our new methods have rather clear advantages over the LS based methods.

Experiment 2. We reconsider the linear model

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_2 X_2 + \varepsilon,$$

which is the same in form as in Experiment 1. However, here the model has the mean-variance-uncertainty as $\varepsilon \sim N([3,5] \times [0,4])$. The other experiment conditions are designed as $X_k \sim N(0,1), k = 1,2,3,$ $m = 10$ and $n = 20$. The values of ε are generated by the following way. Firstly, the values μ_i of the mean and the values σ_i^2 of the variance are generated from the uniform distributions $U[3,5]$ and $U[0,4]$ respectively, and then the values $\varepsilon_{ij}, j = 1, \cdots, n,$ of ε are generated from the common normal distribution $N(\mu_i, \sigma_i^2)$ for $i = 1, \cdots, m$. The simulation results are listed in Table 2. For the MSE of the parameter estimation, the results are similar those in Experiment 1, i.e., the MES of the LS estimation is smaller than that of the new estimation because the new method only uses the data in a small subinterval in principle. However, when the mean-uncertainty and variance-uncertainty appear in the model, both the MPE and the APE of the new one are significantly smaller than those of the LS estimator. Particularly, the prediction by the LS seems to be totally invalid. It indicates that ignoring the model-uncertainty will lead to a serious prediction risk.

Experiment 3. In this experiment, we consider the following high-dimensional

Table 2: Simulation results of estimation and prediction for Experiment 2 with $m = 10$ and $n = 20$

MSE	β_1	β_2	β_2	MPE	APE
$\widetilde{\beta}_G$	0.1258	0.2769	0.2398	14.4254	6.8787
$\hat{\beta}_{LS}$	0.1141	0.1891	0.1879	36.0253	21.2932

linear model

$$Y = \sum_{j=1}^{p} \beta_j X_j + \varepsilon.$$

In the simulation procedure we choose $p = 40$, $\beta_j = 1$ for $1 \leq j \leq 5$ and $\beta_j = 0$ for all $j \geq 6$, $X \sim N_{40}(0, I_{40})$, $\varepsilon \sim N(\{0\} \times [1,4])$ and the sample size satisfies $m = 10$ and $n = 200$. Like the condition in Experiment 1, this model has the mean-certainty. We consider the common LS and G-normal estimation, as well as use the common LASSO and the G-normal LASSO (G-LASSO) defined in Section 4 to select variables and estimate parameters simultaneously. The tuning parameter λ is determined by the CV. Under the above experiment condition, for the common LASSO estimation, the value of λ is chosen as $\lambda_{LS} = 0.0604$; for the G-LASSO, the value of λ is chosen as $\lambda_G = 0.3377$. The simulation results are reported in Table 3 and Figure 1. In Table 3, GNR, LSR, Lasso-GNR and Lasso-LSR stand for the G-normal regression, LS regression, LASSO-G-normal regression and LASSO-LS regression, respectively. The simulation results in Table 3 can verify that the new methods can efficiently reduce maximum prediction error. From Figure 1 we have the following findings: (1) The the LS methods are more stable than the new methods; (2) Like the common LASSO, the G-normal LASSO can efficiently select the active variables.

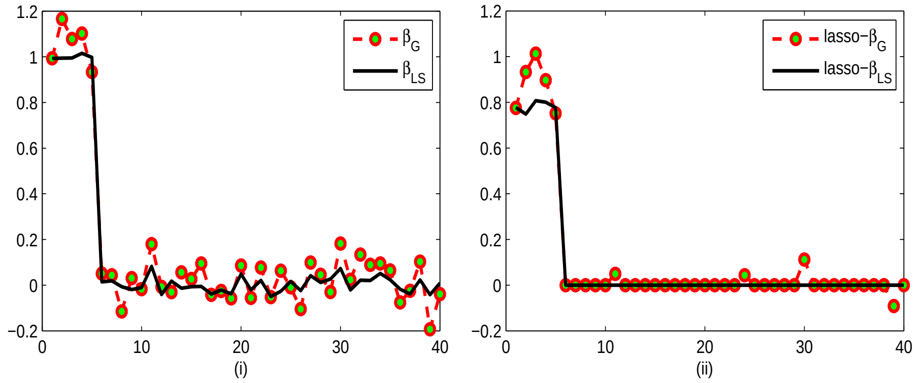


Figure 1: The figures of estimation for Experiment 3 with independent covariates.

Table 3: Simulation results of prediction for Experiment 3 with independent covariates

Models	GNR	LSR	Lasso-GNR	Lasso-LSR
MPE	4.0365	4.5443	4.0221	4.1206
APE	2.5334	2.0801	2.2363	2.0573

To further examine the behaviors, here we consider the correlated covariates: $X \sim N_{40}(0, \Sigma)$, where Σ is 40×40 matrix with the (ij)-element as

$$\Sigma_{ij} = \begin{cases} 1, & \text{for } i = 1, \\ 0.5, & \text{for } i \neq j. \end{cases}$$

The other experiment conditions are designed as the same as the above. The simulation results are presented in Figure 2. The performances of the figures in Figure 2 are similar to those in Figure 1, but they are not as stable as before because of the correlation among the covariates.

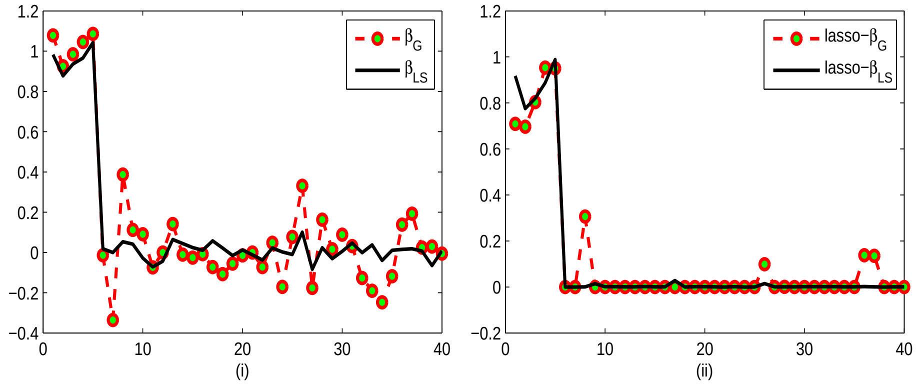


Figure 2: The figures of estimation for Experiment 3 with correlated covariates.

Experiment 4. In this experiment, the model is designed as the same in form as that in Experiment 3, but the model has both the mean-uncertainty and the variance-uncertainty. Formally, the error is distributed as $\varepsilon \sim N([5,10] \times [1,4])$, which has the mean-variance-uncertainty. We first consider the simulations for the GNR and the LSR without use of the LASSO, the results being reported in Figure 3. Figure 3(i) verifies again that the parameter estimation of the LSR is more stable than that of the GNR. On the other hand, Figure 3(ii) provides a clear evidence that with the mean-variance-uncertainty, the LSR has rather large values of the MPE and the APE and therefore the LSR prediction is invalid completely, but the GNR can significantly reduce both the MPE and the APE. These results imply that under the mean-variance-uncertainty framework, ignoring the mean-uncertainty will result in a serious prediction risk, but the new method can efficiently reduce prediction risk by the use of the information of the mean-uncertainty of the error ε.

Now we consider variable selection and parameter estimation by the LASSO. Under the experiment conditions above, we get $\lambda_G = 0.6494$ and $\lambda_{LS} = 0.4670$ via the CV method. For the Lasso-GNR, the simulation re-

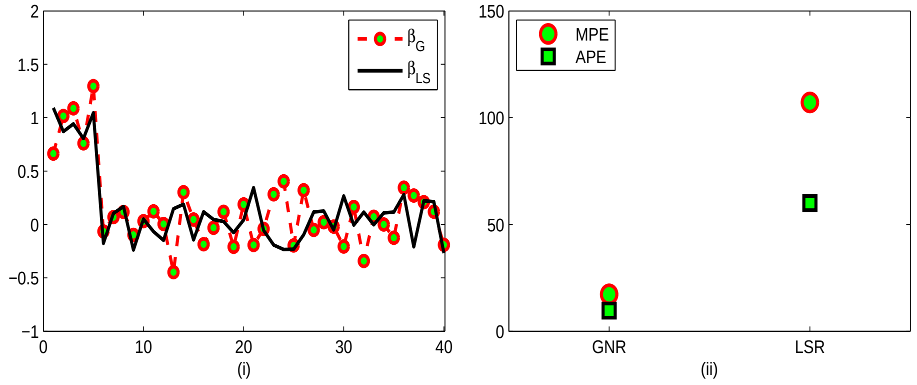


Figure 3: The figures of estimation and prediction for Experiment 4.

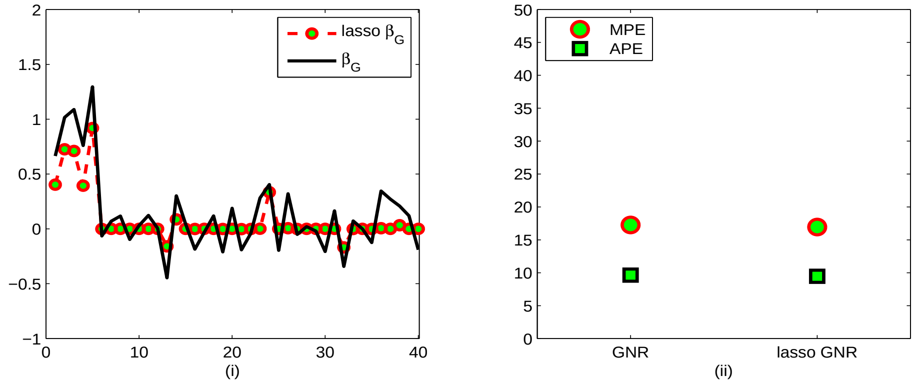


Figure 4: The figures of GNR and Lasso-GNR estimation and prediction for Experiment 4.

sults are given by Figure 4. It shows that the new method can efficiently select active variables and at the same time, the prediction risks are rather small. For the Lasso-LSR, the simulation results are presented in Figure 5. By comparing Figure 5 and Figure 6, we have a clear evidence to show that the new method can obviously reduce the prediction risk, but the LS prediction collapses nearly.

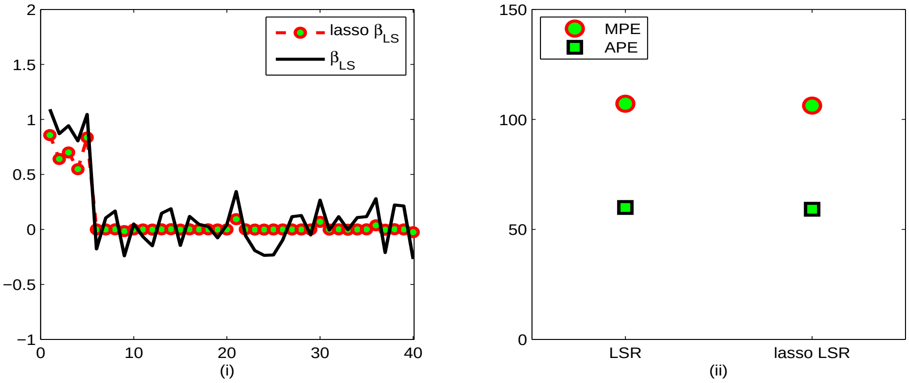


Figure 5: The figures of LSR and Lasso-LSR estimation and prediction for Experiment 4.

In short, our methods are robust to the uncertainties of mean and variance. Particularly, for the case of serious mean-uncertainty, the classical methods may collapse, but our new methods can successfully eliminate the impact of mean-uncertainty and construct efficient prediction. The main disadvantage of the new methods is the instability, more precisely, the resulting estimation has relatively large variance since the mini-max estimation only uses the data in a subinterval, essentially.

5.2 Real data analysis

Non-performing loan (NPL) is always an important object to be monitored in financial market. To investigate the relationship between the NPL ratio and a set of economic indicators, we use our models, together with the new estimation methods, to fit the real data published in Vendors Database of China (2000-2010). We also compare our fittings with the LS fittings that ignore the distribution-uncertainty. According to the indicator system in Vendors Database, after the indicators with which the data are incomplete are deleted, we choose the following indicators as initial choices: loan-deposit Ratio (X_1), capital adequacy ratio (X_2), core capital adequacy ratio (X_3), liquidity ratio of short-term assets of RMB (X_4), liquidity ratio of short-term assets of foreign currencies (X_5), proportion of loans from other banks (X_6), proportion of loans to other banks (X_7), ten largest customers loan ratio (X_8), single biggest customer loan ratio (X_9) and NPL provision coverage (X_{10}). Because the indicators X_j are percentages, they are transformed to $\tilde{X}_j = \log\frac{a_j+X_j}{b_j-X_j}$ for some constants $a_j > 0$ and $b_j > 1$, and then $\tilde{X}_j$ are centralized so that the centralized versions of $\tilde{X}_j$ have zero mean. In the following, we still use X_j to denote the transformed and centralized indicators for simplicity. According the observation time order, the data are decomposed into five sets, in which the numbers of valid data are $n_1 = 26$, $n_2 = 25$, $n_3 = 21$, $n_4 = 20$ and $n_5 = 31$ respectively.

From the real data analyses given below, we will have the following findings: (1) With model-uncertainty technique, the new methods in most cases have more efficient fitting than the LS does; (2) Particularly, when the technique of mean-variance uncertainty is employed to fit the real data, a more precise fitting can be obtained.

5.2.1 Case 1 (Mean-certainty model)

We first use a model with mean-certainty to fit the data.

(1) If the variable selection is not taken into account, by our method of variance-uncertainty, we get an empirical model as

$$\begin{aligned} M_G\text{-}1: \ Y = &-0.2602X_1 + 0.1922X_2 - 0.3953X_3 - 0.2513X_4 + 0.0607X_5 \\ &-0.1808X_6 + 0.0727X_7 + 0.4314X_8 - 0.1503X_9 - 0.5656X_{10}. \end{aligned}$$

With this model, the maximum prediction error and average prediction error have values:

$$MPE(M_G\text{-}1) = 1.6009, \ APE(M_G\text{-}1) = 0.4632.$$

As a contrastive method, the LS is used to build model, the resulting empirical model has the following form:

$$\begin{aligned} M_{LS}\text{-}1: \ Y = &-0.2590X_1 + 0.1843X_2 - 0.3972X_3 - 0.2268X_4 + 0.0543X_5 \\ &-0.2073X_6 + 0.0914X_7 + 0.2734X_8 - 0.0315X_9 - 0.5884X_{10}. \end{aligned}$$

The corresponding prediction errors have the following values:

$$MPE(M_{LS}\text{-}1) = 1.4396, \ APE(M_{LS}\text{-}1) = 0.4323.$$

By comparing the prediction errors, we see that in this case our method has no advantage over the LS fitting. We will analyze the causes in the following studies.

(2) Since some indicators among the ten economic indicators have clear correlation and the number of data is relatively small, the fittings above are inefficient. It is necessary to select variables so that the final model is parsimonious and workable. Now we use the Lasso, together with variance-uncertainty, to build an empirical model, which has the following form:

$$\begin{aligned} M_G\text{-}2: \ Y = &-0.1770X_1 - 0.0111X_2 - 0.1878X_3 \\ &-0.0549X_4 + 0.1397X_8 - 0.5956X_{10}. \end{aligned}$$

By this treatment, the prediction errors have the following values:

$$MPE(M_G\text{-}2) = 0.8443, \ APE(M_G\text{-}2) = 0.3500.$$

By use of the Lasso, the inactive predictors are removed from the model, the model size is significantly reduced and prediction effectiveness is improved clearly.

If variance-uncertainty is ignored, the Lasso-LS empirical model has the following form:

$$\begin{aligned} M_{LS}\text{-}2: \ Y = &-0.0387X_1 - 0.0269X_2 - 0.0542X_3 \\ &+0.0722X_8 + 0.0352X_9 - 0.5381X_{10}, \end{aligned}$$

and the corresponding prediction error have the following values:

$$MPE(M_{LS}\text{-}2) = 1.0420,\ APE(M_{LS}\text{-}2) = 0.4258.$$

By comparing M_G-2 and M_{LS}-2, we have a clear evidence that our method can reduce prediction errors.

5.2.2 Case 2 (Mean-variance-uncertainty model)

We can verify that $\underline{\mu} = -0.1833$ and $\overline{\mu} = 0.1747$. Thus, such a mean-uncertainty is not ignorable. To improve data fitting, both mean-uncertainty and variance-uncertainty are taken into account in the following modeling procedure.

(1) Without use of variable selection, the model with mean-variance-uncertainty has the following empirical expression:

$$\begin{aligned}\widetilde{M}_G\text{-}1:\ Y = &-0.2315X_1 + 0.1888X_2 - 0.4765X_3 - 0.2673X_4 + 0.0129X_5 - 0.2590X_6\\ &+0.0798X_7 + 0.6331X_8 - 0.3093X_9 - 0.5374X_{10} + 0.1747.\end{aligned}$$

This model leads to the prediction errors as

$$\widetilde{MPE}(M_G\text{-}1) = 0.9182,\ \widetilde{APE}(M_G\text{-}1) = 0.3837.$$

Comparing $\widetilde{M}_G$-1 with both M_G-1 and M_{LS}-1, the model $\widetilde{M}_G$-1 has the following two distinctive features: it uses a relatively large value to predict the NPL ratio and the prediction errors are significantly reduced.

(2) By use of the Lasso, the model with mean-variance-uncertainty has the following empirical expression

$$\widetilde{M}_G\text{-}2:\ Y = -0.0389X_1 - 0.0420X_2 - 0.1309X_3 - 0.5108X_{10} + 0.1747.$$

By this treatment, the prediction errors are reduced to

$$MPE(M_G\text{-}2) = 0.7305,\ APE(M_G\text{-}2) = 0.4362.$$

This model may be the best one among all the models mentioned above because it has both the smallest model size and the smallest MPE.

In short, a flexible model that has mean-variance-uncertainty can relatively precisely fit the real data and is parsimonious and workable.

6 Appendix

6.1 Definition of sublinear expectation

Let Ω be a given set and $\mathcal{H}$ be a linear space of real valued functions defined on Ω. Suppose that $\mathbb{E} : \mathcal{H} \to \mathbb{R}$ satisfies the following properties: for all $U, V \in \mathcal{H}$,

(i) Monotonicity: If $U \geq V$ then $\mathbb{E}[U] \geq \mathbb{E}[V]$;

(ii) Constant preservation: $\mathbb{E}[c] = c$ for any constant c;

(iii) Sub-additivity: $\mathbb{E}[U + V] \leq \mathbb{E}[U] + \mathbb{E}[V]$;

(iv) Positive homogeneity: $\mathbb{E}[\lambda U] = \lambda \mathbb{E}[U]$ for each $\lambda \geq 0$.

Then $(\Omega, \mathcal{H}, \mathbb{E})$ is called a sublinear expectation space.

It can be verified that (iii) and (iv) together imply

(v) Convexity:

$$\mathbb{E}[\alpha U + (1-\alpha)V] \leq \alpha \mathbb{E}[U] + (1-\alpha)\mathbb{E}[V] \ \text{ for } \alpha \in [0, 1].$$

Furthermore, (ii) and (iii) together lead to

(vi) Cash translatability:

$$\mathbb{E}[U + c] = \mathbb{E}[U] + c \ \text{ for any constant } c.$$

6.2 Proofs

Proof of Proposition 2.1 We only need to prove the second result. It is clear that (2.5) yields

$$\mathbf{x}\mathbb{E}[Y|\mathbf{x}] = \mathbf{x}\mathbf{x}'\beta.$$

Note that the distribution of $\mathbf{x}$ is certain. Consequently,

$$E\{\mathbf{x}\mathbb{E}[Y|\mathbf{x}]\} = E[\mathbf{x}\mathbf{x}']\beta.$$

This implies the second result of the proposition. □

Proof of Proposition 2.2 We only need to prove the second result. It is obvious that by (2.5) we have

$$\mathbf{x}\mathbb{E}[Y|\mathbf{x}] = \mathbf{x}\mathbf{x}'\beta + \overline{\mu}\mathbf{x}$$

and consequently

$$E\{\mathbf{x}\mathbb{E}[Y|\mathbf{x}]\} = E(\mathbf{x}\mathbf{x}')\beta + \overline{\mu}E[\mathbf{x}].$$

This implies the second result of the proposition. □

Proof of Theorem 3.1 It follows from *C1* that

$$\frac{1}{n}\sum_{j=1}^{n}\varepsilon_{ij}^2 = \sigma_i^2 + \delta_n,$$

where δ_n is of order $O_p(1/n)$ and is free of β. Consequently,

$$\max_{1\leq i\leq m}\frac{1}{n}\sum_{j=1}^{n}\varepsilon_{ij}^2 = \sigma_{i_*}^2 + \delta_n.$$

Denoted by β^0 the true value of β. Then

$$\begin{aligned}
&\max_{1\leq i\leq m}\frac{1}{n}\sum_{j=1}^{n}[Y_{ij} - \beta'\mathbf{x}_{ij}]^2 \\
&= \max_{1\leq i\leq m}\frac{1}{n}\sum_{j=1}^{n}\left[\varepsilon_{ij} - (\beta-\beta^0)'\mathbf{x}_{ij}\right]^2 \\
&= \max_{1\leq i\leq m}\frac{1}{n}\sum_{j=1}^{n}\left[\varepsilon_{ij}^2 - 2(\beta-\beta^0)'\mathbf{x}_{ij}\varepsilon_{ij} + (\beta-\beta^0)'\mathbf{x}_{ij}\mathbf{x}_{ij}'(\beta-\beta^0)\right].
\end{aligned}$$

Note that $\mathbf{x}_{ij}, i = 1, \cdots, m, j = 1, \cdots, n$, are independent and identically distributed with zero mean and variance σ_i^2. By comparing the asymptotic orders of every terms in the above expression, we have

$$\max_{1 \le i \le m} \frac{1}{n} \sum_{j=1}^{n} [Y_{ij} - \beta' \mathbf{x}_{ij}]^2$$

$$= \frac{1}{n} \sum_{j=1}^{n} \left[\varepsilon_{i_* j}^2 - 2(\beta - \beta^0)' \mathbf{x}_{i_* j} \varepsilon_{i_* j} + (\beta - \beta^0)' \mathbf{x}_{i_* j} \mathbf{x}_{i_* j}' (\beta - \beta^0)\right] + \delta_n.$$

As was shown that $\varepsilon_{i_* j}$ and δ_n are independent of β. Thus, to get the estimator of β, minimizing $\max_{1 \le i \le m} \frac{1}{n} \sum_{j=1}^{n} [Y_{ij} - \beta' \mathbf{x}_{ij}]^2$ is equivalent to minimizing

$$\sum_{j=1}^{n} \left[-2(\beta - \beta^0)' \mathbf{x}_{i_* j} \varepsilon_{i_* j} + (\beta - \beta^0)' \mathbf{x}_{i_* j} \mathbf{x}_{i_* j}' (\beta - \beta^0)\right].$$

We rewrite the above objective function as

$$Z_n(\gamma) = \sum_{j=1}^{n} \left[-2 \frac{\varepsilon_{ij} \mathbf{x}_{i_* j}'}{\sqrt{n}} \gamma + \gamma' \frac{\mathbf{x}_{i_* j} \mathbf{x}_{i_* j}'}{n} \gamma\right].$$

The function $Z_n(\gamma)$ is obviously convex and is minimized at $\hat{\gamma}_n = \sqrt{n}(\hat{\beta}_G - \beta^0)$. It follows from the Lindeberg-Feller central limit theorem and *C1* that

$$Z_n(\gamma) \xrightarrow{d} Z_0(\gamma) = -2W'\gamma + \gamma' E(\mathbf{x}\mathbf{x}')\gamma,$$

where $W \sim N(0, \sigma_{i_*}^2 E(\mathbf{x}\mathbf{x}'))$. The convexity of the limiting objective function, $Z_0(\gamma)$, assures the uniqueness of the minimizer and consequently, that

$$\sqrt{n}(\hat{\beta}_G - \beta) = \hat{\gamma}_n = \arg\min \tilde{Z}_n(\gamma) \xrightarrow{d} \hat{\gamma}_0 = \arg\min Z_0(\gamma).$$

(See, e.g., Pollard 1991, Hjørt and Pollard 1993, Knight 1998). Finally, we see $\hat{\gamma}_0 = (E(\mathbf{x}\mathbf{x}'))^{-1} W$ and the result follows. □

Proof of Theorem 3.2 The definitions of the two estimations lead directly to the conclusions of the theorem. □

Proof of Theorem 3.3 From the proof of Theorem 3.1 we see that $\hat{\beta}_G$ is actually the common LS estimator of β obtained by data $(Y_{i_*j}, \mathbf{x}_{i_*j}), j = 1, \cdots, n$. Thus $\hat{\beta}_G = \beta + O_p(1/n)$, where β is the true regression coefficient given by (2.6) in the mean-certainty model. When $E[\mathbf{x}] = 0$, the true regression coefficients in the mean-certainty model and the mean-uncertainty model are the same as given in (2.6) and (2.9). Moreover, by the the same argument as used in the proof of Theorem 3.1, we have

$$\max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n}[Y_{ij} - \beta'\mathbf{x}_{ij}] = \max_{1\le i\le m} E_{F_i}[\varepsilon] + O_p(1/n).$$

The above discussion ensures that

$$\hat{\overline{\mu}} = \max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n}[Y_{ij} - \beta'\mathbf{x}_{ij}]+O_p(1/n) = \max_{1\le i\le m} E_{F_i}[\varepsilon]+O_p(1/n) = \overline{\mu}+O_p(1/n),$$

where F_i is the distribution of data in I_i. Consequently,

$$\max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \beta'\mathbf{x}_{ij} - \hat{\overline{\mu}}\right]^2 = \max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n}[Y_{ij} - \beta'\mathbf{x}_{ij} - \overline{\mu}]^2 + O_p(1/n).$$

On the other hand,

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n}[Y_{ij} - \beta'\mathbf{x}_{ij} - \overline{\mu}]^2 = \frac{1}{n}\sum_{j=1}^{n}[\varepsilon_{ij} - \mu_i - (\overline{\mu} - \mu_i)]^2\\ &= \sigma_i^2 + (\overline{\mu} - \mu_i)^2 + O_p(1/n).\end{aligned}$$

Then,

$$\max_{1\le i\le m} \frac{1}{n}\sum_{j=1}^{n}[Y_{ij} - \beta'\mathbf{x}_{ij} - \overline{\mu}]^2 = \frac{1}{n}\sum_{j=1}^{n}[Y_{k_*j} - \beta'\mathbf{x}_{k_*j} - \overline{\mu}]^2 + O_p(1/n).$$

By the above result, $E[\mathbf{x}] = 0$ and the same argument as used in the proof of Theorem 3.1, we can prove the theorem. □

Proof of Theorem 3.4 The proof of the theorem follows directly from the definitions of the two estimators. □

References

Artzner, Ph., Delbaen, F., Eber, J. M. and Heath, D. (1997). Thinking coherently. *RISK*, **10**, 86-71.

Bradley, R. C. and Bryc, W. (1985). Multilinear forms and measures of dependence between random variables. *J. Multivariate Anal.*, **16**, 335-367.

Briand, Ph., Coquet, F., Hu, Y., Mémin J. and Peng, S. (2000). A converse comparison theorem for BSDEs and related properties of g-expectations. *Electron. Comm. Probab*, **5**, 101-117.

Candés, E. J. and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.* **35**, 2313-2351.

Chen, Z. and Epstein, L. (2002). Ambiguity, risk and asset returns in continuous time. *Econometrica*, **70**(4), 1403-1443.

Chen, Z. and Peng, S. (2000). A general downcrossing inequality for gmartingales. *Statist. Probab. Lett.*, **46**(2), 169-175.

Coquet, F., Hu, Y., Mémin J. and Peng, S. (2002). Filtration-consistent nonlinear expectations and related g-expectations. *Probab. Theory Relat. Fields*, **123**, 1-27.

Denis, L. and Martini, C. (2006). A theoretical framework for the pricing of contingent claims in the presence of model uncertainty. *The Ann. of Appl. Probability*, **16**(2), 827-852.

Denis, L., Hu, M. and Peng S. (2011). Function spaces and capacity related to a sublinear expectation: application to G-Brownian motion pathes. *Potential Anal.*, **34**, 139-161.

El Karoui, N., Peng, S. and Quenez, M.C. (1997). Backward stochastic differential equation in finance. *Mathematical Finance*, **7**(1): 1-71.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J Amer Statist Assoc*, **96**, 1348-1360.

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.*, **32**, 928-961.

Föllmer, H. and Schied, A. (2004). *Statistic Finance, An introduction in discrete time* (2nd Edition), Walter de Gruyter.

Gao, F. Q. (2009). Pathwise properties and homeomorphic flows for stochastic differential equations driven by G-Brownianmotion. *Stochastic Processes and their Applications*, **119**, 3356-3382.

Huber,P. J. (1981). *Robust Statistics*, John Wiley & Sons.

Hjørt, N. and D. Pollard (1993), Asymptotics for minimizers of convex processes. *Statistical Research Report.*

Kolmogorov, A. N. and Rozanov, U. A. (1960). On the strong mixing conditions of a stationary Gaussian process. *Probab. Theory Appl.*, **2**, 222-227.

Knight, K. (1989). Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics*, **17**, 261-278.

Li, X and Peng, S. (2011). Stopping times and related Itô's calculus with G-Brownian motion. *Stochastic Processes and their Applications*, **121**, 1492-1508.

Lu, C. R. and Lin, Z. Y. (1997). Limit theories for mixing dependent variables. Science Press, Beijing.

Peng, S. (1997). Backward SDE and related g-expectations, in Backward Stochastic Differential Equations, Pitman Research Notes in Math. Series, No.364, El Karoui Mazliak edit. 141-159.

Peng, S. (1999). Monotonic limit theorem of BSDE and nonlinear decomposition theorem of Doob-Meyers type. *Prob. Theory Rel. Fields*, **113**(4), 473-499.

Peng, S. (2004). Filtration consistent nonlinear expectations and evaluations of contingent claims. *Acta Mathematicae Applicatae Sinica.* English Series **20**(2), 1-24.

Peng, S. (2005). Nonlinear expectations and nonlinear Markov chains, *Chin. Ann. Math.*, **26**B(2), 159-184.

Peng, S. (2006). G-Expectation, G-Brownian Motion and Related Stochastic Calculus of Itôs type, The Abel Symposium 2005, Abel Symposia 2, Edit. Benth et. al., 541-567, Springer-Verlag, 2006.

Peng, S. (2008). Multi-dimensional G-Brownian motion and related stochastic calculus under G-expectation. *Stochastic Processes and their Applications*, **118**(12), 2223-2253.

Peng, S. (2009). Survey on normal distributions, central limit theorem, Brownian motion and the related stochastic calculus under sublinear expectations. *Science in China Series* A: Mathematics, **52**, 7, 1391-1411.

Pollard, D. (1991). Asymptotics for least absolute deviation regression Estimators. *Econometric Theory*, **7**, 186-199.

Rosazza, G. E. (2006). Some examples of risk measures via g-expectations. *Insurance: Mathematics and Economics*, **39**, 19-34.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci.*, **42**, 43-47.

Rosenblatt, M. (1970). Density estimates and Markov sequences, in *Nonparametric Techniques in Statistical inference*, ed. M. Puri, London: Cambridge University Press, pp. 199-210.

Soner, M., Touzi, N. and Zhang, J. (2011a). Martingale representation theorem under G-expectation. *Stochastic Processes and their Applications*, **121**, 265-287.

Soner M, Touzi N, Zhang J. (2011b) Quasi-sure stochastic analysis through aggregation. *Electronic Journal of Probability*, **16**, 1844-1879.

Soner M, Touzi N, Zhang J. (2012). Wellposedness of second order backward SDEs. *Probability Theory and Related Fields*, **153**, 149-190.

Soner M, Touzi N, Zhang J. (2013). Dual formulation of second order target problems. *Annals of Applied Probability*, **23**, 308-347.

Song, Y.(2012). Uniqueness of the representation for G-martingales with finite variation. *Electron. J.Probab.*, **17**, 1-15.

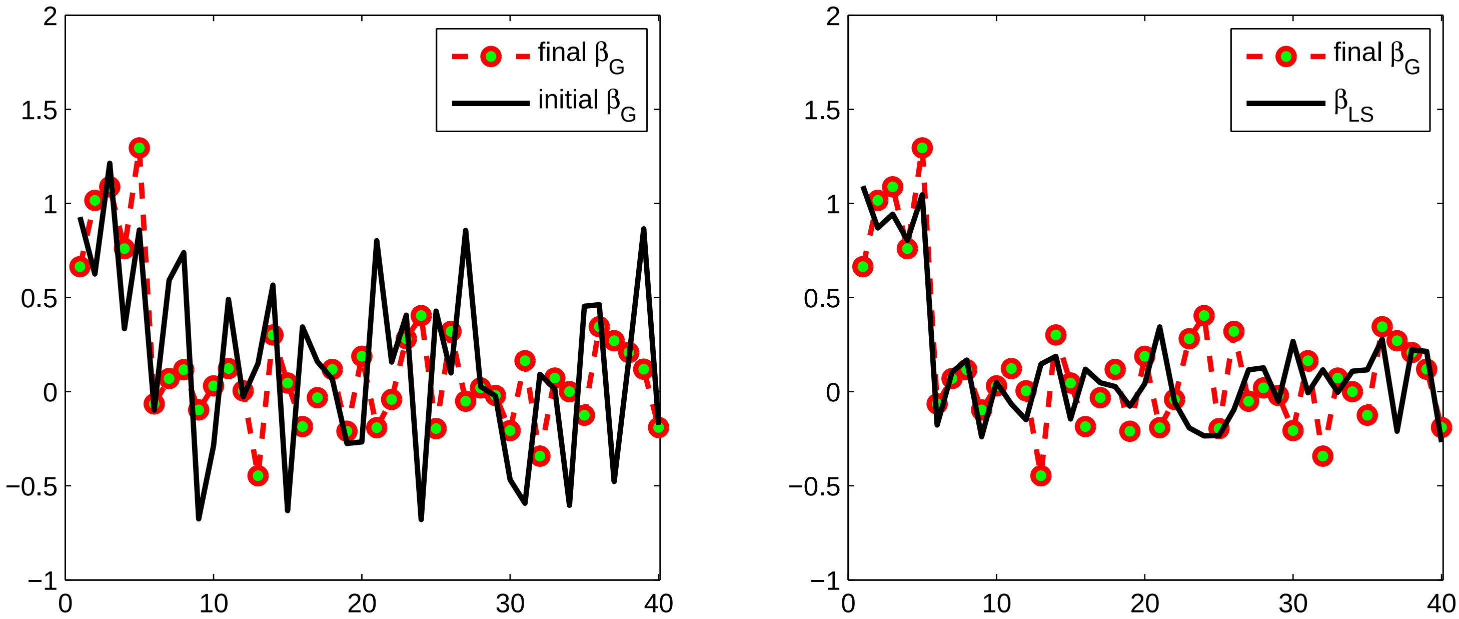
Strassen, V. (1964). Messfehler und information, *Z. Wahrsheinlichkeitstheorie verw. Gebiete*, **2**, 267-284.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B*, **58**, 267-288.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, New York.

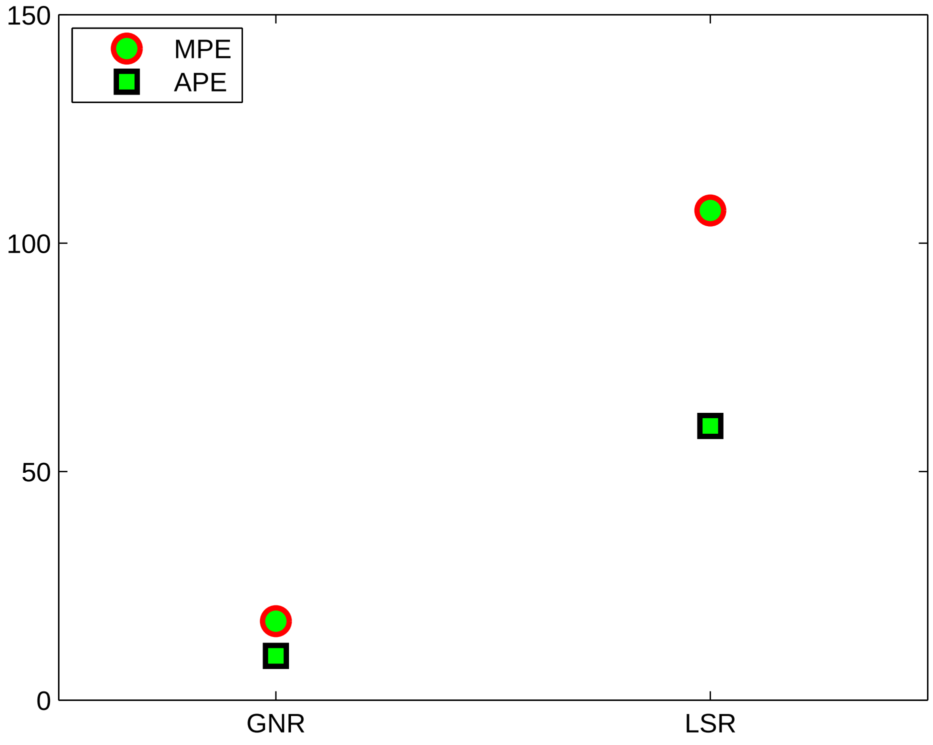
Xu, J. and Zhang, B. (2009). Martingale characterization of G-Brownian motion. *Stochastic Processes and their Applications*, **119**, 232-248.

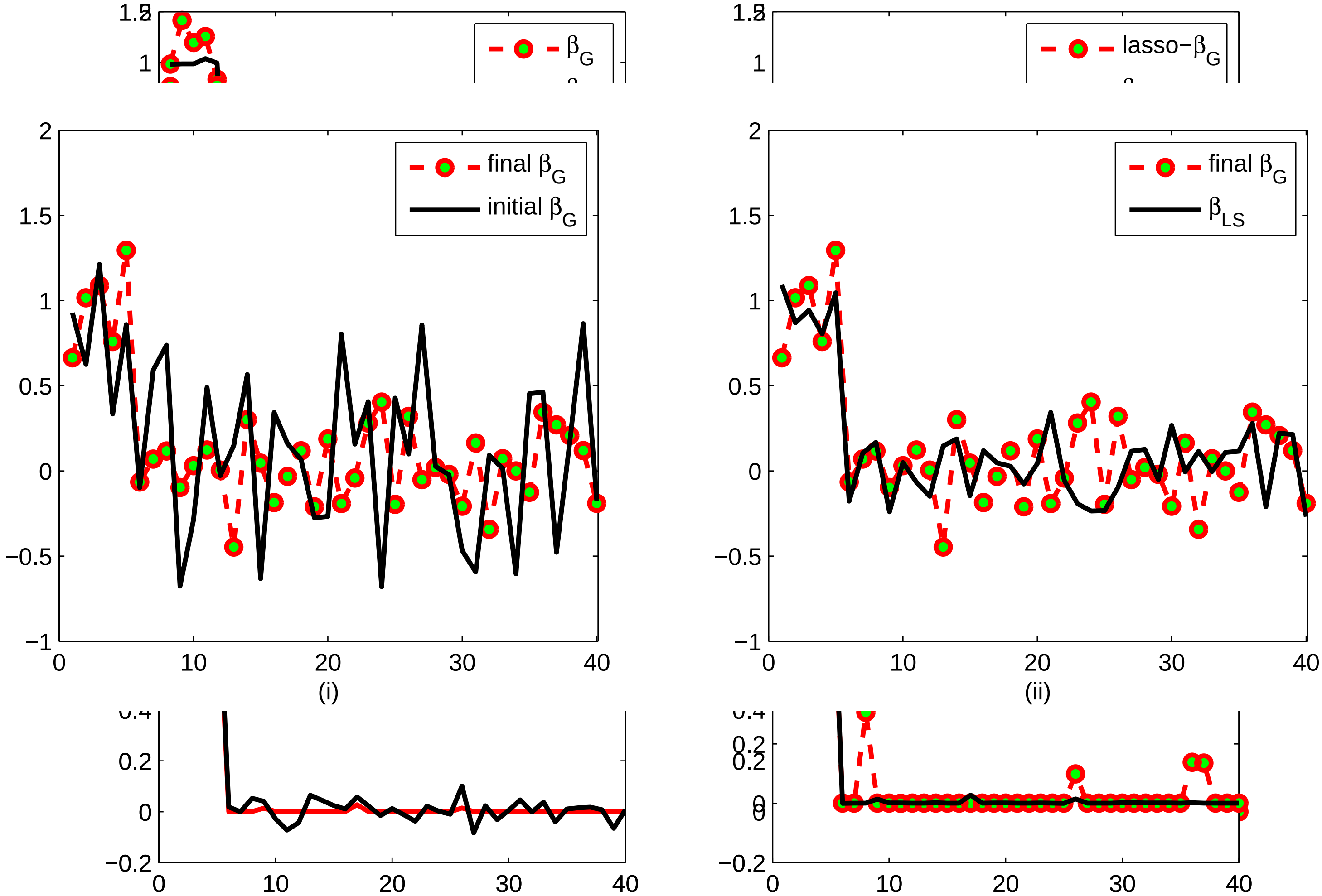
Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, **101**, 1148-1429.


2
1.5
1
0.5
0
−0.5
−1
0
10
20
30
40
final β_G
initial β_G
2
1.5
1
0.5
0
−0.5
−1
0
10
20
30
40
final β_G
β_{LS}

(i)

(ii)

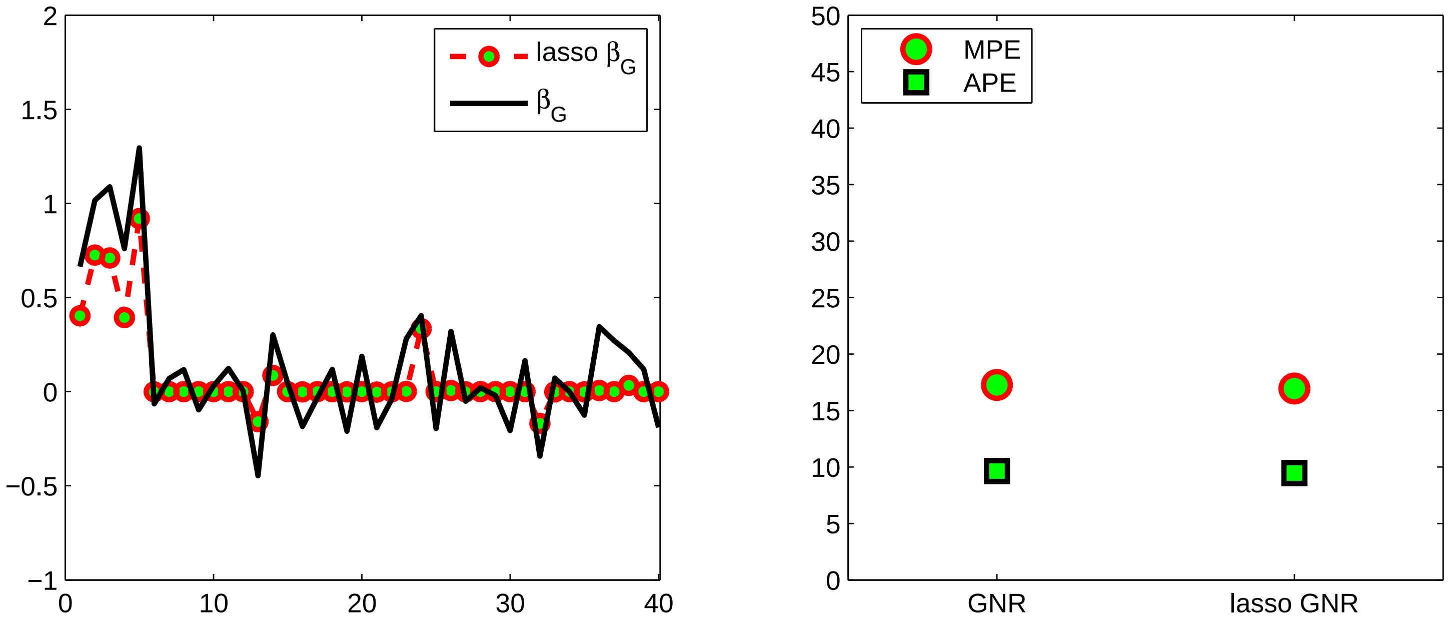

150
100
50
0
MPE
APE
GNR
LSR



final β_G
initial β_G
final β_G
β_{LS}
β_G
lasso−β_G
2
1.5
1
0.5
0
−0.5
−1
0
10
20
30
40
0.4
0.2
−0.2

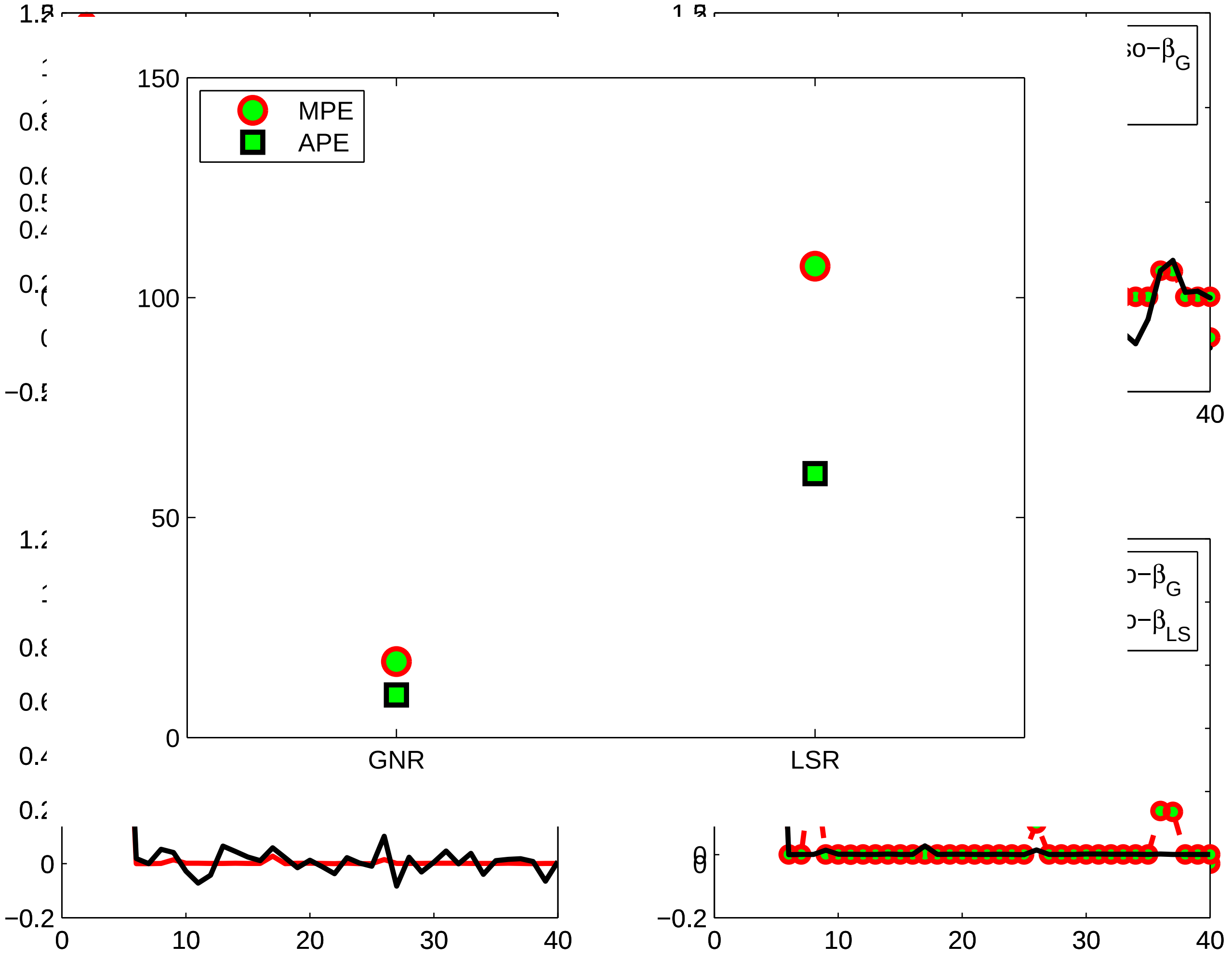
(i)

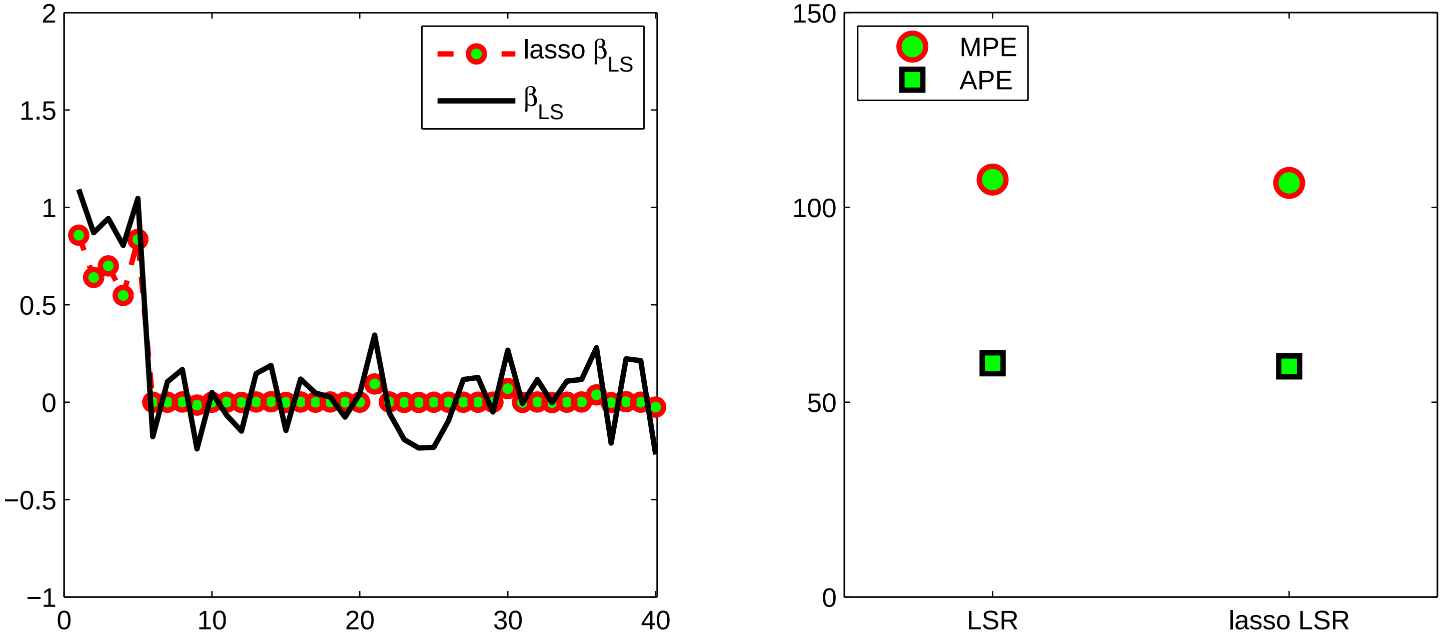
(ii)


lasso β_G
β_G
2
1.5
1
0.5
0
−0.5
−1
0
10
20
30
40
MPE
APE
50
45
40
35
30
25
20
15
10
5
0
GNR
lasso GNR

(i)

(ii)


150
100
50
0
GNR
LSR
MPE
APE
so−β_G
o−β_G
o−β_{LS}
1.2
0.8
0.6
0.4
0.2
0
−0.2
−0.5
0
10
20
30
40


2
1.5
1
0.5
0
−0.5
−1
0
10
20
30
40
lasso β_{LS}
β_{LS}
150
100
50
0
MPE
APE
LSR
lasso LSR

(i)

(ii)